Attorney Docket No.: A02001

Express Mailing Label No.: EF193968052US

PERFORMING RETIMING EFFECTS ON SYNCHRONIZED DATA IN AN EDITING SYSTEM

By

Randy M. Fayan
31 Orchard Street, #1
Medford, Massachusetts 02155

Katherine H. Cornog
26 Chestnut Street
Newburyport, Massachusetts 01950

Both citizens of the United States of America

PERFORMING RETIMING EFFECTS ON SYNCHRONIZED DATA IN AN EDITING SYSTEM

BACKGROUND

During editing of an audiovisual work, it is common to expand or to contract an audiovisual clip to fit a particular time slot or to create a perceptual rate change. Such operations may be performed for a creative purpose or for a technical purpose. There are a number of terms used to describe such operations, including "motion effect," "fit-to-fill," "time scaling," or, generally, a "retiming effect." If the operation has the effect of changing the perceived playback rate from one constant rate to another constant rate, then the operation may be called a "constant" retiming effect. If the operation has the effect of providing a perceived playback rate that varies over the length of the clip, then the operation may be called a "rampable" retiming effect. The variable rate change of a rampable retiming effect commonly is defined by a function curve. The function curve typically is a speed curve that describes how the perceived playback speed varies from the original playback speed, in terms of a percentage of the original playback speed.

There are many ways in which such retiming effects may be implemented. Commonly, a retiming effect is performed by stretching one of the audio or video tracks to match timing of the other track. Video and audio also typically are retimed separately. Video is retimed in video post-production using techniques such as sample and hold, blended frames or motion-based interpolation. In audio post-production, the audio may be replaced with another soundtrack, may remain unchanged or may be retimed using techniques such as time scaling or pitch shifting. Both audio retiming techniques change the perceived playback rate of the audio, but time scaling may be used to avoid or control modifying the pitch.

Because the video and audio typically are retimed separately, the video and audio typically are retimed using different speed curves or using the same function curve but sampled by different sampling rates. Either technique makes it difficult to retain synchronization between the audio and video.

SUMMARY

During editing of an audiovisual work, it would be desirable to see and hear the result of a rampable retiming effect on a clip of synchronized audio and video data that produces a retimed result of synchronized audio and video data. An editing system that processes such rampable retiming effects retimes and synchronizes playback of both the audio and video data in the clip.

To allow synchronized playback, a retiming function that defines the rampable retiming effect is used to generate new audio and video samples at appropriate output times. In particular, for each output time, a corresponding input time is determined from the output time by using the retiming function. The retiming function may be a speed curve, a position curve that maps output times to input times directly or a mapping defining correspondence times between points in the video data and points in the audio data. An output sample is computed for the output time based on at least the data in the neighborhood of the corresponding input time, using a resampling function for the type of media data. The neighborhood of the corresponding input time is a plurality of samples from points in time surrounding the input time. The number of input samples actually used depends on the resampling function used. A resampling function generates an output sample from a plurality of input samples at different points in time by combining information from the plurality of input samples. An example resampling function is interpolation. Synchronization is achieved by ensuring that the input times determined to correspond to output times for video samples correspond to the input times determined to correspond to the same output times for audio samples. In other words, synchronization is achieved by using the same mapping of input and output times based on the retiming function to drive the resampling functions.

There are several ways to perform such retiming in several different workflows for creating an audiovisual work that includes a rampable retiming effect. The typical workflows include using a video editing system to edit the audiovisual work, with a focus on the video, followed by using an audio editing system to edit, primarily, the corresponding audio. For some audiovisual works, such as music videos, the audio is edited first and the video is retimed to match the audio. In some workflows, audio is edited on the video editing system. In each of these applications, the retiming function

may be a speed curve, position curve or a mapping defining correspondence times between points in the video data and points in the audio data.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram of a timeline representing an audiovisual work.

Fig. 2 is a diagram of a timeline representing the audiovisual work with a retimed clip.

Fig. 3 is a diagram of an example speed curve.

Fig. 4A is a diagram of position curves generated from the example speed curve using different step sizes for audio and video data.

Fig. 4B is a diagram of a position curve generated from the example speed curve using the same step sizes for audio and video data.

Fig. 5 is an illustration of a graphical user interface that allows a user to define a retiming effect.

Fig. 6 is an illustration of a graphical user interface that allows input audio and video times to be related to an output time.

Fig. 7 is a diagram of position curves resulting from the user interface of Fig. 6.

Fig. 8 is a block diagram of a system that applies a retiming function to audio and video data.

Fig. 9 is a flow chart describing how a retiming function may be applied using a video editing system for editing the audio.

Fig. 10 is a flow chart describing how a retiming function may be applied using a video editing system and an audio editing system.

DETAILED DESCRIPTION

Fig. 1 illustrates a part of a typical user interface of a video editing system for editing audiovisual programs, called a timeline 10. A timeline represents the program using one or more video tracks 12 and one or more audio tracks 14. In this example, one video track and two audio tracks are shown. Various interfaces, other than a timeline with tracks shown in Fig. 1, may be used to define the audiovisual program in a video

editing system, such as interfaces common to compositing systems, audio editing systems and animation tools. On each track in timeline 10, a sequence of clips 16 is shown. In this example, the clips designated with an "X" (18) are synchronized, and start and end at the same time. The "X" need not be part of the displayed timeline. Thus, this example represents a video clip synchronized with a two-track audio clip. A clip on the timeline may be defined using a reference to a file that stores actual media data for the clip and a range within the media data stored in that file. Audio and video data may be stored in the same file or in separate files.

The application of a retiming effect to a clip also may, or may not, affect the duration of the clip. In particular, if a clip has a number of samples at a sampling rate, and if that number of samples is processed using the retiming function, then the result of the retiming effect is a new number of samples at the original sampling rate, but with a different perceived playback rate. The clip may become shorter or longer depending on the retiming function. If the clip is in an audiovisual work when the retiming effect is applied, and if the retiming effect modifies the duration of the clip, then the synchronization of the rest of the audiovisual work may be affected. If the retiming effect does not modify the duration, then the synchronization of the rest of the audiovisual work will not be affected.

Accordingly, before a clip is placed in an audiovisual work, a retiming effect may be applied to the specified range of a media file. The duration resulting from the application of the retiming effect to the specified range of the media file defines the duration of the clip that may be placed on a timeline representing the audiovisual work.

After a clip is placed in an audiovisual work, when the retiming effect is applied, the specified range of the media file that is used may be changed so that the retimed clip has the same duration as the original clip. Such a change to the specified range of the media file might exceed the contents of the media file. In such a case, the duration of the retimed clip might be changed, or an error might be signaled to the editor, or the last sample from the media file might be held for the remaining duration of the retimed clip. If the retimed clip includes media that the editor does not want, then the editor could trim the retimed clip.

Referring now to Fig. 2, the original clip 18 (Fig. 1) in the audiovisual program has been retimed to have a perceived playback rate that is faster than the original playback rate. In this example, the clip has a new duration because the number of output samples has changed, as indicated by the reduced length of the clip 28 as shown in the timeline 10.

A retiming effect is specified by a retiming function, which will now be described. To maintain synchronization while retiming, the input times determined for output times for video samples are calculated to correspond to input times determined for the same output times for the audio samples. An input time determined for each output time for video samples may be identical to or offset from an input time determined for the same output time for audio samples. This correspondence can be achieved in several ways.

One way to achieve this correspondence is to use a position curve that maps each output sample time to an input time. The same position curve is used to retime both the audio and the video data. A position curve can be defined in many ways. For example, a user may define the position curve through a user interface that allows a function curve to be defined.

Another way to achieve this correspondence is to use a speed curve that represents the perceived speed of playback, in comparison to the original speed of playback. For example, a speed curve is as shown in Fig. 3. The curve in Fig. 3 represents the speed on the vertical axis and time on the horizontal axis. With the example curve that is shown, playback starts at 50% of the original playback speed, gradually increases to 100% of original playback speed, then gradually decreases to 50% of the original playback speed. A speed curve can be defined in many ways. For example, a user may define the speed curve through a user interface that allows a function curve to be defined.

If a speed curve defines the retiming function, the speed curve is converted to a position curve by computing an integral of the speed curve. The position curve is used to retime the clip. To generate the integral for a curve, digital computers often calculate the integral using numerical methods which approximate a Riemann sum, i.e. summing the

area of the rectangles or trapezoids under the curve using a small value, called the step size, that defines the width of the rectangle or trapezoid. Normally, video editing systems calculate the integral using a step size that equals one over the field rate (e.g., the step size =1/60 sec), but audio editing systems calculate an integral using a step size that equals one over the audio sample rate (e.g., step size = 1/44,100 sec). A constant also is used in the calculation of an integral. This constant may be specified as the time code of the first frame of the source material. Alternatively, the user may specify some other frame of the source material as an "anchor frame" to provide the constant time value for the integral calculation.

Using the speed curve from Fig. 3, the integral for the audio using a step size for a typical audio system is shown at 40 in Fig. 4A. The integral for the video using a typical step size for a video system results in the position curve shown at 42 in Fig. 4A. Notably, the position curves are different. When the position curves are different as in Fig. 4A, the same points in time in the original audio data and the original video data do not map to the same output time. Therefore, if the audio were retimed on an audio system and if the video were retimed on a video system, the resulting retimed data would not be synchronized.

Instead, the position curve is calculated by integrating the speed curve using a step size that is less than or equal to the minimum of the reciprocal of the audio sample rate and the reciprocal of the video sample rate. The integral used to obtain the position curve for video thus is computed using a step size that is less than or equal to a step size corresponding to a sampling rate of the audio data, rather than the reciprocal of the video rate. As a result, using the speed curve of Fig. 3 as an example, the same position curve is used to retime both the audio and video, as indicated at 50 in Fig. 4B. The sampling rate of the audio data used by an audio editing workstation is thus shared with the video editing system.

An example graphical user interface that allows a user to specify a speed curve or position curve for a retiming effect will now be described in connection with Fig. 5. This user interface may be made available to an editor if an editor selects a retiming effect to be applied either to source material or to a clip on the timeline.

In this interface 500, a speed graph 502 and position graph 504 are shown. The editor, using an appropriate selection mechanism, may select either graph for editing. Another selection mechanism can be provided to allow only one or both of the graphs to be shown.

In the speed graph, the vertical axis 508 represents the speed as a percentage of the original playback speed of the clip. The horizontal axis 510 represents the output times. The editor may specify keyframes (e.g., 512) on the graph to associate an output time with a perceived playback speed. A keyframe signifies that, at the output time indicated by the keyframe, the associated perceived playback speed will be applied. For output times between the specified keyframes, the remainder of points on the function curve is interpolated. For output times before the first keyframe and after the last keyframe, values may be extrapolated or may be held constant. Controls may be provided to allow a user to select a keyframe and to navigate among the keyframes. On the speed graph, left and right boundaries, shown as vertical lines 514 and 516, indicate the start and stop times of the retimed clip, thus its duration.

The position graph display 504 shows a representation of an integral of the speed curve. If the speed graph is being edited, then the position graph display shows the corresponding position graph, with corresponding keyframes. The position graph may be edited, in which case the speed graph display shows the corresponding speed graph, with corresponding keyframes.

In the position graph display 504, the vertical axis 522 illustrates input times, in terms of the video time code of the source material to which the retiming effect is being applied. The horizontal axis 524 illustrates output times. Left and right boundaries, shown as vertical lines 526 and 528, represent the start and stop times of the retimed clip, and thus its duration. The top horizontal line 530 represents the end of the source material available in the referenced media file. The bottom horizontal line 532 represents the beginning of the source material available in the referenced media file. The top and bottom horizontal lines allow an editor to determine whether enough source material is available to be processed by the specified retiming effect.

The interface also may allow a user to specify various parameters of the retiming effect, such as the kind of resampling function used. An anchor frame may be selected

by a user in many ways, such as by input of a timecode or by manipulation of a specified keyframe that is designated as an anchor frame. The format of the output data may be specified, if different from the format of the input data. In such a case, the resampling function also converts the format of the data. The interface also may provide status information, such as the duration of the effect.

Another way to achieve the correspondence of input times and output times is to use different position curves for the audio and video data so long as the input times of audio and video events that should be synchronized map to the same output time. In particular, video and audio events are identified and related to each other and a corresponding output time. Such a mapping provides separate position curves for the audio data and the video data. Because the resulting position curves directly describe the mapping of output times to input times, the input audio times and the input video times are obtained directly from the position curves. Because the position curves map video and audio events to the same output times, synchronization of those events is retained.

Visual and audio events may be identified manually or automatically. Examples of audio and video events are phonemes and visemes (visual appearance of a mouth articulating a phoneme). An example of corresponding phonemes and visemes is an onset of a 'P' sound and lips closing prior to pronunciation of a 'P'. Identification of corresponding audio and video events can be achieved by an editor or by a computer, and can be tracked by placing locators on a timeline.

A user interface that allows different position curves to be defined for audio and video will now be described in connection with Figs. 6 and 7. Referring to Fig. 6, three parallel timelines are shown: a video timeline 60, an audio timeline 62 and an output timeline 64. Each event is indicated by three locators, one for each of the video, audio and output times for the event. Each locator for an event has a common shape. Each event has a different shape. The time of a locator in the output timeline may correspond to a locator in the video track, to one in the audio track or to a time different from either. In Fig. 6, three events are shown. In the first event noted at 66, the audio is retimed to match the original video. The video event time and the output time are the same. In the second event noted at 68, the video is retimed to match the original audio. The audio

event time and the output time are the same. In the third event noted at 61, both the video and the audio are retimed to produce a new output pacing.

From the relative position of the locators in the video, audio and output tracks, two position curves are derived separately. To define the curve for an input track, for each event in the output track a point or keyframe is defined, where x is the output time of the event on the output track and y is the input time of the event on the input track. The set of points or keyframes obtained for the set of events may then be joined by a function, for example, a piecewise linear function or a smooth function such as a spline or a Bezier curve, that passes through all of the points.

Position curves corresponding to the events shown in Fig. 6 are shown in Fig. 7. The first curve maps output times to input times for retiming video, at 72, and the other curve maps output times to input times for retiming audio, at 70. The two curves are plotted on the same axes for convenience and comparison. The curves shown are piecewise linear, but also may be splines or Bezier curves to provide smooth motion or pace changes. Synchronization is achieved because corresponding video and audio events map to a single output time.

An example application of the technique shown in Figs. 6 and 7 is in automatic dialog replacement (ADR). ADR is used, for example, if the audio track of a shot is poor and is replaced with a new recording. To maintain lip sync, the timing of events in the new recording must match the timing of events in the video. Using the above technique for specifying a position curve, both the timing and the pacing of events may be changed from that of any of the original recordings while maintaining lip sync.

After an editor specifies a retiming function, the specification may be stored as data associated with the specified clip of video and audio data. Preferably, the retiming function is defined as a position curve using a continuous or piecewise continuous function to allow different systems with different sampling rates to use the same position curve. If a speed curve is used, a step size value used to generate a retimed clip also may be stored to allow different systems to generate the same position curves from the speed curve.

In a video editing system in which clips are defined by references to files, the retiming function may be applied to the media data in the data files during playback or before playback. If the retiming function is applied before playback, the original media data files may remain unchanged and the clip may be modified to include a reference to a data file that stores the retimed media data, which may be called a rendered or precompute file.

Having now described ways to specify the retiming function as a speed curve or position curve, and how the retiming function is associated with a clip, how synchronized audio and video is retimed will now be described.

In general, for each output time for an audio sample, a corresponding input time is determined from the retiming function. An output audio sample is computed at the output time based on at least the audio data in the neighborhood of the corresponding input time using a resampling function. The neighborhood is a plurality of audio samples from points in time surrounding the input time. The number of input audio samples actually used depends on the resampling function that is used to compute the audio output sample. An audio resampling function generates an output audio sample from a plurality of input audio samples at different points in time by combining information from the plurality of input audio samples.

For each output time for a video sample, a corresponding input time is determined from the output time and the retiming function, such that an input time determined for an output time for a video sample corresponds to an input time determined for the same output time for an audio sample. An output video sample is computed at the output time based on at least the video data in the neighborhood of the corresponding input time using a resampling function. The neighborhood is a specified number of video samples from points in time surrounding the input time. The number of input video samples so specified depends on the resampling function that is used to compute the output video sample. A video resampling function generates an output video sample from a plurality of input video samples at different points in time by combining information from the plurality of input video samples.

The output audio samples may be computed using any of a number of resampling functions, including, but not limited to time scaling and pitch shifting. Techniques for time scaling and pitch shifting are described for example in "A Sines+Transients+Noise Audio Representation for Data Compression and Time/Pitch Scale Modifications," by Scott Levine and Julius Smith III, in Proceedings of the 105th Audio Engineering Society Convention, San Francisco, 1998. A commercially available product that performs time scaling is called SPEED, available from Wave Mechanics, of Burlington, Vermont.

The output video samples may be computed using any of a number of resampling functions, including, but not limited to blended frames or motion-based interpolation. Motion based interpolation is described in U.S. Patent Application Serial No. 09/657,699, entitled "INTERPOLATION OF A SEQUENCE OF IMAGES USING MOTION ANALYSIS."

A retiming effect may be rendered to create a new media file containing the retimed clip, or may be applied in real time to source material. Whether a retiming effect may be processed in real time during playback depends on the processor speed of the computer, the disk access speed and the resampling technique used for the video and the audio and whether any degradation of the image quality, in either the spatial or temporal dimensions, is permitted.

Referring now to Fig. 8, a block diagram of a system for performing such retiming will now be described. The inputs used in generating synchronized retiming of an audiovisual clip are an audio stream 80, a video stream 82, and either a user-defined speed curve 84 or position curve 86. A system may provide the capability of an editor to specify either a speed curve, position curve or both. If the position curve is used, the time values used of the audio samples and video data may be obtained directly from the function. If a speed curve is used it is integrated, as shown at 88, to produce a position curve. The integral used to convert the speed to position is performed according to the specification above - in particular, the position curve is computed by integrating the speed curve using a step size that is less than or equal to the reciprocal of the audio sample rate. The position curve is used by both video retiming 81 and audio retiming 83 to produce the corresponding retimed video stream 85 and retimed audio stream 87.

Referring to Fig. 9, a flowchart representing operation of a video editing system that produces retimed audio and video data of an audiovisual work using such a retiming effect will now be described. In such a system a retiming effect on a clip of synchronized audio data and video data is performed to produce a retimed clip of synchronized audio and video data in the audiovisual work. The system associates (90) a definition of a retiming function for a rampable retiming effect that maps output times to input times with the clip of synchronized audio data and video data. The synchronized audio data and video data is processed according to the retiming function to produce the retimed clip. Such processing includes, for each output time for an audio sample, determining (92) a corresponding input time from the output time and the retiming function, and computing an output audio sample at the output time based on at least the audio data in the neighborhood of the corresponding input time. Similarly, for each output time for a video sample, a corresponding input time is determined (94) from the output time and the retiming function, such that the input time determined for the output time for a video sample corresponds to the input time determined for the same output time for an audio sample. An output video sample is computed at the output time based on at least the video data in the neighborhood of the corresponding input time. The retimed clip is placed (96) in the audiovisual work. The retimed clip also can be played back.

Referring to Fig. 10, a dataflow diagram of a system using a video editing system 100 and an audio editing system 102 will now be described. A retiming effect 108 on a clip of synchronized audio data 104 and video data 106 is defined. The video editing system 100 and an audio editing system 102 enable an editor to see and hear the retimed clip as part of an audiovisual work. The specifications of the work are transferred between the video editing system 100 and the audio editing system 102 to allow modification to the video or to the audio. Such a situation may arise, for example, where different individuals are working in a group on different parts of the audiovisual work.

In one embodiment, the video editing system 100 provides retimed video data. To produce the retimed video data, if the retiming function is defined as a speed curve, the video editing system also receives a description of the sampling rate of the audio from the audio editing system. The audio editing system 102 produces the retimed audio data from the original audio data and synchronizes the retimed audio data with the retimed

video data. In particular, the audio editing system receives a definition of a retiming function, the original audio data and the audiovisual work including a retimed video clip. The audio editing system then processes the audio data according to the retiming function to produce a retimed audio clip. As a result, the retimed audio is synchronized with the retimed video in the audiovisual work.

In another embodiment, the audio editing system 102 provides retimed audio data, and the video editing system 100 produces the retimed video data from the original video data and synchronizes the retimed video data with the retimed audio data. In particular, the video editing system receives a definition of a retiming function, the video data and the audiovisual work including a retimed audio clip. In the exchange of information between the audio editing system 102 and the video editing system 100, the sampling rates used for the audio data is shared if the retiming function is defined as a speed graph. The video data is then processed according to the retiming function to produce a retimed video clip. As a result, the retimed video is synchronized with the retimed audio in the audiovisual work.

Having now described an example embodiment, it should be apparent to those skilled in the art that the foregoing is merely illustrative and not limiting, having been presented by way of example only. Numerous modifications and other embodiments are within the scope of one of ordinary skill in the art. For example, the example above describes synchronization of audio and video data. Retiming of synchronized data streams in this manner can be extended to any temporal data streams, or data streams with temporal relationships, where the data in the different data streams is of different types. For example, an event such as a trigger in an interactive television program is tied to a particular output time or input time of another media type, such information can be used to adjust the timing of the trigger if the other media is retimed.

These and other modifications and embodiments are contemplated as falling within the scope of the invention.

What is claimed is: